DOCUMENT RESUME

ED 453 274 TM 032 811

AUTHOR Pomplun, Mark
TITLE Cluster Analysis: A Method To Develop School Level Normative Score Profiles To Support School Improvement?
PUB DATE 2001-04-14
NOTE 22p.; Paper presented at the Annual Meeting of the American Educational Research Association (Seattle, WA, April 10-14, 2001).
PUB TYPE Reports - Research (143) -- Speeches/Meeting Papers (150)
EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Cluster Analysis; *Educational Improvement; Elementary Secondary Education; Mathematics; *Norms; *Profiles; Reading; *Schools; *Scores

ABSTRACT

This study investigated cluster analysis as a method to identify common school score profiles. Scores from more than 100,000 students were used to create school level subscores in reading and mathematics. Four scores for constructing and extending meaning in fiction and nonfiction reading passages were analyzed in reading. Seven subscores were analyzed in mathematics. For reading, the resulting profiles grouped about 70% of the students into a consistently flat (high, average, or low) profile. For mathematics, the resulting profiles grouped only about 7% of the students into a consistently flat profile. Of the other profiles, six had one low or high score, seven had two low subscores, and one had three low scores. While the reading profiles later had clear instructional implications, the profiles from mathematics were less straightforward. More research should be done to validate the identified profiles and their usefulness in instruction. (Author/SLD)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 453 274

Cluster analysis: A method to develop school level normative score profiles to support school improvement?

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

M. Pomplun

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

1

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

Mark Pomplun

Senior Research Scientist

Riverside Publishing

425 Spring Lake Drive

Itasca, IL 60143-2079

Tele 630-467-6056

Fax 630-467-7126

mark_pomplun@hmco.com

TM032811

BEST COPY AVAILABLE

Abstract

This study investigated cluster analysis as a method to identify common school score profiles. Scores from over 100,000 students were used to create school level subscores in reading and mathematics. Four scores for constructing and extending meaning in fiction and nonfiction reading passages were analyzed in reading. Seven subscores were analyzed in mathematics. For reading, the resulting profiles grouped about 70% of the students into a consistently flat (high, average or low) profile. For mathematics, the resulting profiles grouped only about 7% of the students into a consistently flat profile. In the other profiles, six profiles had one low or high score, seven had two low subscores, and one had three low scores. While the reading profiles had clear instructional implications, the profiles from mathematics were less straightforward. More research needs to be done to validate the identified profiles and their usefulness in instruction.

Cluster analysis: A method to develop school level

normative score profiles to support school improvement?

A critical step in the process of school improvement is identification of curricular areas that need improvement. Assessments to support school improvement usually provide a total score and subscores for each subject area. After the total score is used to determine a schools' overall functioning, school personnel usually use the subscore results to identify strengths and weaknesses. However, the interpretation of patterns of subscores has been a long standing source of controversy in educational measurement. Many practictioners interpret the results of subscores from large-scale achievement assessments even though most methodologists view this practice as problematic. In the past, univariate base rates and statistical significance testing of pairs of subscores have been used to perform formal profile analysis. However, neither of these approaches accounted for correlations among subscores, allowed multivariate testing or were developed for the analysis of sets of scores.

Recently, studies have appeared that use a normative multivariate approach to perform profile analysis that considers correlations among subscores and allows analysis of sets of scores. In 1994, Roid studied patterns of subscores from a large-scale writing assessment using the six trait model. He used cluster analysis to identify subgroups of students who showed similar patterns of trait scores. In 1999, Konold, Glutting, McDermott, Kush and Watkins also used cluster analysis to identify subgroups of students with similar score patterns on the *Wechsler Intelligence Scale for Children-Third Edition*. They identified the most prevalent patterns and the percentage of students who

displayed each pattern. These studies suggest that cluster analysis can detect differences in the level and shape of subscore profiles. The present study explored the usefulness of cluster analysis for the identification of profiles to help educators identify areas at the school level that need remediation strategies. Two sets of subscores were used, one with relatively higher intercorrelations and one set with relatively lower intercorrelations.

Cluster analysis: Cluster analysis is a family of procedures that can be used to combine schools into clusters or groups based on the similarity of the patterns of their scores (Subhash, 1996). The end result of the procedure is the formation of groups of schools which are similar to each other with respect to the scores of interest and different from schools in the other groups. In contrast to other methods like discriminant analysis and factor analysis that could be used to identify the score profiles that categorize schools into groups, cluster analysis makes no assumptions about the probability distribution from which the sample was obtained and the calculations are simple and straightforward.

Cluster analysis has five steps (Subhash, 1996). The first is the selection of the measure used to judge the similarity or distance of two cases. These include squared Euclidean distances and correlation coefficents. The second step is the decision to use either hierarchical or nonhierarchical clustering. Generally, in hierarchical clustering, a wider variety of similarity measures and more ways to calculate differences between cases are available while in nonhierarchical clustering the user is limited to one distance measure and one method to calculate distances between groups. In addition, hierarchical analysis allows the user to evaluate a wider range of possible cluster solutions while nonhierarchical requires the user to specify the number of clusters beforehand. However, the nonhierarchical procedure can handle many more cases than the hierarchical analysis.

Most authors recommend using both methods in the same analysis with the hierarchical method used first to determine the number of clusters and then followed by a nonhierarchical analysis to refine the definition of the clusters and to classify all the subjects (Aldenderfer & Blashfield, 1984; Subhash, 1996).

The third step is selection of the method to calculate the distance or similiarity between the groups of cases to be clustered. These include the centroid, average-linkage, and Ward's method. Although the same distance measure would be used, in the centroid method the similarity is calculated between the average scores for each pair of schools or clusters, in the average-linkage method similarities are calculated for each pair of cases then averaged, and in Ward's method, between the cases within each cluster. The fourth step or choice is the method used to determine the number of clusters that best describe the data. Although several statistics are available for this task, their sampling distributions are unknown, and as a result, the statistics are basically heuristics (Subhash, 1996). The last step in the cluster analysis is the validation and interpretation of the clusters. This can involve cluster relationships with external variables, examination of fit statistics, and the consistency of the derived clusters with theoretical expectations.

The present study used cluster analysis to identify common school score profiles from two third grade tests of reading and mathematics. Hierarchical cluster analyses were used to identify these profiles and then a nonhierarchical method was used to classify the schools into these profiles.

Method

Subjects: This study used results from large-scale reading and mathematics achievement tests used to support a state assessment in a Midwest state. At the third grade over 60,000

students participated in the reading assessment and over 58,000 in the mathematics assessment. Only schools with more than nine students were included in the analysis. When aggregated at the building level, 895 buildings had third grade reading scores and 877 buildings had third grade mathematics scores.

Instrument and Scores. The program had tests in reading, mathematics, writing, science and social studies at several grade levels. The objectives measured by the tests at each grade were parallel to those specified in the state currriculum guidelines. Items were both multiple choice and constructed response. Multiple-choice items were scored 0 or 1 and constructed-response items were scores from 0 to 4 with the scale varying across the items. At the third grade, there were four reading subscores with constructing/examining meaning and examining/extending meaning scores associated with a fiction/poetry section and a nonfiction passage. The correlations at the school level between these reading subscores ranged from .90 to .92. At the third grade, mathematics had seven subscores including patterns/functions, problem-solving strategies, number relations, geometry, measurement, estimation, and data analysis/probability. The correlations at the school level between these subscores ranged from .68 to .92.

Analyses: The analyses for both studies were conducted in two stages. Scores were first standardized into z-scores to have all subscores on the same scale and then each school's mean score was subtracted from it's subscores to control for level or elevation effects. In the first stage, exploratory hierarchical cluster analysis along with correlational analysis were conducted to identify initial profiles. These profiles were then used in the second stage of nonhierarchical cluster analysis (Roid, 1994).

The first stage began with an inspection of correlations between the subscores. Then, subscores were input into a hierarchical cluster analysis (SPSS, 9.0). In the hierarchical cluster analysis, the average linkage method with Pearson correlations as similarity measures was employed. This combination was used because of the generally positive intercorrelations among the subscores in the study and a desire to detect patterns of differentiated scores (Roid, 1994). The average-linkage method uses the average distance between pairs of subscores, one pair for each of the subscores, as the distance between two schools or clusters.

To determine the number of clusters for each test, the number of clusters was graphed against the correlation between the two schools or clusters merged to form the last or newest cluster (Aldenderfer & Blashfield, 1984). This test is a variant of the scree test used in factor analysis. A marked flattening in the graph (or an "elbow") suggests that relatively different schools or clusters were merged in the last cluster.

The clusters identified in the above analyses were then used in the second stage of nonhierarchical cluster analysis to classify the full sample of schools. Specifically, the clusters identified in the first stage were used as cluster centered mean scores to perform the nonhierarchical clustering conducted with K-Means cluster analysis (SPSS, 1999). This K-Means method uses squared Euclidean distances as the distance measure.

Results

Table 1 displays descriptive statistics for the scores used in this study. The number of schools, the maximum number of points, mean, standard deviation, percent correct and reliability are shown for the scores from each of the tests. For reading, the percent corrects suggest that items which addressed fiction passages, extending or

constructing meaning, were relatively higher or the easier part, and the items which addressed nonfiction passages were relatively lower or the more difficult part. For mathematics, the percent corrects suggest that items which measured patterns/algebra and geometry were the easier parts, and the items which addressed measurement were the more difficult part. The statistics in Table 1 also indicate that these subscores from both reading and mathematics had reliabilities at or above .90 which is usually considered excellent score reliability.

Tables 2 and 3 show the correlations at the school level among the subscores from each test. These correlations in Table 2 suggest that the four reading scores were measuring skills which overlapped to a large degree. For example, if the correlations were corrected to adjust for the less than perfect reliabilities shown in Table 1 the corrected correlations would be close to one. As a result, flat score profiles of consistent scores were expected for many schools. The correlations in Table 3 indicate that many schools will not have consistent mathematics score profiles. While, subscores for patterns/algebra, problems, and measurement all correlate .90 or above with each other, geometry subscores to a large extent and estimation subscores to a lesser extent had lower correlations with the other subscores. This suggests that at least some profiles will reflect high and low subscores in these two areas.

Next, hierarchical cluster analyses were performed on the subscores from each test to identify score profiles that would be useful in classifying and describing schools. Before these analyses, all subscores were standardized to account for the slightly different number of items for some subscores on some forms and to decrease the influence of different score variances. In addition, to partial out the influence of overall

score level on the profiles, the mean z-score for each school was subtracted from that school's z-score subscores. The number of clusters from the hierarchical analyses were evaluated using the scree test. The profiles identified in the hierarchical cluster analysis along with the flat profiles suggested in the correlational analysis were input into the second stage of the analysis, nonhierarchical cluster analysis.

The profiles identified in the first stage were used as cluster centroids for the nonhierarchical analysis operationalized through the iterative K-Means procedure (SPSS, 1999). The results of this cluster analysis are shown for reading in Tables 4 and 5 and for mathematics in Tables 6 and 7. Tables 4 and 6 contain a verbal description of each profile, and the z-score mean centroids for each score in each profile. Tables 5 and 7 display the percentage of all the schools classified in each profile, a total average scale score, the average number of students in the profile schools (a measure of school size) and a fit index for the cases in each profile category. The fit index shows the average distance of the scores for the schools classified in each group from the centroids for that group.

Table 4 shows the mean z-scores for each score in each reading profile and the verbal label for the profile group. The results in Table 5 address the validation and interpretation of the score profiles. For reading, the "Flat" profile describes 73% of the schools, or the largest group. The "Low Fiction - Constructing" profile describes 6% or the next largest. The "Low Fiction" and "Low Nonfiction" profiles describe 3% each or the smallest groups. Interestingly, the 'Flat' profile group has the most students per schools, the smallest standard deviation among scores, the highest total score and the lowest or best fit. The groups with the fewest schools 'Low Extending', 'Low

Nonfiction' and 'Low Fiction' generally have the smallest schools (fewest number of average students), more variation among their scores within each school, lower overall total scores and somewhat lower fit.

Table 6 shows the mean z-scores for each mathematics profile and the verbal label for the profile group. Table 7 shows that for mathematics, the "High Geometry" and "Low Geometry" profiles describe 15% and 13% of the schools respectively, or the largest groups. The "Flat" profile describes 7% or the next largest group. The "Low Estimate" profile describes 3% or the smallest group. Interestingly, the overall mean scores and the fit statistics do not appear to have the clear relationship these indices had in the reading analysis. In Table 7, the standard deviation or the variability of the scores within schools had a high correlation with the fit index, with high variability associated with poorer fit and low variability related to better fit. There was also a tendency for profile groups with more students per school to have smaller variability among their scores. The "Lo Geometry" group had the highest mean score (221), the second largest number of schools, the highest standard deviation, the poorest fit and fifth largest in school size. However, the 'Flat' group had the third largest group of schools, the largest size of schools but the lowest variability of scores and best fit.

To further examine these relationships, correlations were calculated at the school level among number of students per school, standard deviation of subscores, total overall score, and fit index. Table 8 displays the correlation for the schools with the reading profiles above the diagonal and schools with the mathematics profiles below the diagonal. For both reading and mathematics, the largest correlations are between the standard deviation and the fit index. Schools with larger variation among their scores

have larger misfit. Schools with smaller variation among their scores have smaller misfit. In mathematics, this relationship is almost perfect. The relationship between standard deviation, number of students per school, and misfit is moderate in both areas ranging from -.291 to -.413. The correlation between total score and the other indices is the lowest ranging from -.108 to .102.

These results appear to be related to the use of a fit statistic based on the distance between each school's scores and the mean scores for it's cluster profile. Apparently, because larger schools have more stable scores and smaller schools have less stable scores, larger schools will tend to have better fit or have more of their scores closer to those cluster means and smaller schools will tend to have poorer fit or have one or more scores farther from the cluster means. At the school level, there is a small probably practically insignificant relationship between the total score and the other indices.

To further examine the fit of the reading and mathematics profiles, schools with relatively large misfit were examined individually. An examination of the subscore profiles for these schools revealed that most misfit was due to either one or two extreme subscores or unusual profiles like low Extending Fiction and Constructing Nonfiction subscores.

Summary

The results from this study indicate that cluster analysis can be useful for classifying schools based on subscore profiles. In reading, the high correlations among the subscores supported the 73% of schools classified as "Flat" or as having consistent scores across the four areas. The other 28% of the schools were classified into 8 profiles that were theoretically consistent with the structure of the test. These profiles had clear

instructional implications. For mathematics, the correlations for some subscores were relatively lower than those for reading. As a result, the mathematics subscores showed more subscore profiles and many fewer schools classified as consistent or "Flat". Only 7% of the schools were classifed as "Flat" or as consistently low, average or high. Six of the 15 profiles had one area of weakness or strength, seven had two areas that were relative weaknesses, and one had three areas of weakness. The geometry and estimation subscores had the lowest correlations with the other subscores and were involved in the definition of relatively more profiles (5 each) than the other subscores. Ten of the fifteen profiles were defined at least partially by one of these two subscores. As a result, most schools were classified primarily by their performance on geometry and/or estimation.

The results also showed that some smaller schools and schools with unusual score profiles had relatively large misfit values. In a large-scale testing program, these misfitting schools would either not be classified or classified in an additional category clearly labeled as unclassified. The classification of smaller schools needs further attention because they have more normal variation in their scores both between subscores and in terms of extreme single scores, The fit measure used in this study summed the distance from each school's subscores and the means for all the schools in the cluster. As a result, schools with larger standard deviations usually had more misfit and were more frequently smaller schools. The values used in this study for the cluster centroids or means were chosen after hierarchical cluster analysis grouped schools with similar subscores together. This may have resulted in profiles with scores that were not statistically different for some of the smaller schools.

This study used a two-step process which included hierarchical and nonhierarchical clustering. The hierarchical step was more useful when there were more subscores, subscores with lower intercorrelations, and subscores that were not directly connected or organized theoretically. Both more subscores and subscores with lower correlations increases the possible number of profiles. Hierarchical analysis helps identify the most common. The nonhierarchical step was able to quickly classify the nearly 1,000 schools in each analysis. However, the nonhierarchical analysis only allowed the use of one measure of fit. This index for each school was based on the summed distances of each subscore from the cluster means. As a result, the fit was not reflective of the overall pattern of scores and was influenced by extreme subscores. A better fit index would reflect the fit of the full profile with one possibility fitting polynomial curves that measured the overall pattern of the subscores.

By identifying patterns of score profiles, cluster analysis provides a foundation for the interpretation of school profiles. However, because this was an exploratory study, more research needs to be conducted to validate the obtained profiles. The next step in this process would be to examine the instructional and curricular characteristics of schools with each profile to identify factors contributing to the common profiles.

References

Aldenderfer, A. S., & Blashfield, R. K. (1984). *Cluster Analysis.* Newbury Park: Sage Publications.

Gustafsson, J., & Snow, R. E. (1997). Ability profiles. In R. F. Dillon (Ed.), *Handbook on Testing*, (pp. 107-135) Westport, Connecticut: Greenwood Press.

Konold, T. R., Glutting, J., J., McDermott, P. A., Kush, J. C., & Watkins, M., M. (1999). Structure and diagnostic benefits of a normative subtest taxonomy developed from the WISC-III standardization sample. *Journal of School Psychology, 37* (1), 29-18.

Roid, G. (1994). Patterns of writing skills derived from cluster analysis of direct-writing assessments. *Applied Measurement in Education, 7* (2), 159-170.

SPSS, Inc. (1999). *SPSS 9.0.* Chicago: SPSS.

Subhash, S. (1996). Applied Multivariate Techniques. New York: John Wiley & Sons, Inc.

Table 1. School Level Descriptive Statistics for Third Grade Tests

Test/Score	Maximum Score	Mean	Standard Deviation	Percent Correct	School * Reliability
Reading N=895 Fiction					
Construct Meaning	9	6.22	.95	69	.93
Extend Meaning	17	11.47	1.38	67	.94
Nonfiction Construct Meaning	10	5.95	1.11	60	.93
Extend Meaning	10	6.62	.99	66	.95
Math N=877					
Patterns	5	3.12	.60	62	.93
Problem	8	4.24	1.00	53	.94
Number	11	5.89	1.13	54	.95
Geometry	5	3.16	.39	63	.90
Measure	9	4.65	1.10	52	.94
Estimate	4	2.23	.49	56	.92
Data	6	3.32	.61	55	.91

* For Reading, reliability is for average building with 67 students and mathematics for average building with 65 students. Score reliability will be lower for buildings with fewer students and higher for buildings with more students.

Table 2. Correlations between School Level Reading Subscores

Reading	Fiction Construct Meaning	Fiction Extend Meaning	Nonfiction Construct Meaning	Nonfiction Extend Meaning
Fiction Construct Meaning	1.00	.90	.91	.90
Extend Meaning		1.00	.92	.92
Nonfiction Construct Meaning			1.00	.92
Extend Meaning				1.00

Table 3. Correlations between School Level Mathematics Subscores

Mathematics	Pattern	Problem	Number	Geometry	Measure	Estimate	Data
Pattern	1.00	.92	.90	.76	.90	.85	.89
Problem		1.00	.91	.77	.91	.86	.90
Number			1.00	.75	.92	.87	.89
Geometry				1.00	.75	.68	.75
Measure					1.00	.89	.91
Estimate						1.00	.85
Data							1.00

Pattern = patterns/functions, Problem = problem solving strategies, Number = number relations, Measure = measurement, Estimate = estimation, Data = data analysis/ probability

Table 4. Descriptions and Score Statistics for School Profiles for Reading Subscores

Profile Description	Fiction Construct Meaning	Fiction Extend Meaning	Nonfiction Construct Meaning	Nonfiction Extend Meaning
Low Extending	.43	-.32	.19	-.31
Low Constructing	-.33	.34	-.35	.34
Low Nonfiction	.38	.27	-.30	-.35
Low Fiction	-.38	-.27	.30	.35
Flat	.00	.00	.00	.00
Low Fiction Constructing	-.52	.18	.18	.16
Low Fiction Extending	.22	-.50	.139	.14
Low Nonfiction Constructing	.20	.11	-.48	.17
Low Nonfiction Extending	.13	.20	.18	-.51

Table 5. Descriptions and Statistics for School Profiles for Reading Subscores

Profile Description	Percent of Schools	Average # of Students	Standard Deviation of Scores	Total Score	Fit
Low Extending	2	49	.42	217	.53
Low Constructing	3	45	.42	215	.46
Low Nonfiction	2	39	.44	217	.56
Low Fiction	2	41	.41	218	.46
Flat	73	69	.20	219	.34
Low Fiction Constructing	6	58	.39	219	.45
Low Fiction Extending	5	50	.38	219	.45
Low Nonfiction Constructing	3	54	.38	217	.48
Low Nonfiction Extending	4	51	.39	220	.45

Table 6. Descriptions and Score Means for Mathematics Subscore Profiles

Profile Description	Pattern	Problem	Number	Geometry	Measure	Estimate	Data
Hi Geometry	-.08	-.05	-.14	.77	-.18	-.18	-.14
Lo Geometry	.15	.12	.07	-.79	.15	.14	.16
Lo Estimate	-.05	.06	.22	.30	-.04	-.51	.02
Lo Patterns Algebra	-.41	-.14	-.04	.13	.00	.33	.13
Lo Data / Geometry	.13	.00	.33	-.46	.02	.21	-.23
Lo Patterns /Geometry	-.35	-.07	.22	-.49	.23	.37	.08
Lo Data / Patterns	-.25	.01	.21	.24	.07	-.02	-.25
Lo Number /Measure	.08	.08	-.26	.02	-.32	.24	.16
Lo Number / Estimate	.08	.26	-.29	.00	-.02	-.26	.23
Lo Number /Estim/Prob	-.02	-.25	-.18	.33	-.01	-.28	.42
Lo Data	.32	-.07	.05	.23	-.06	.00	-.47
Lo Data / Number	.04	-.11	-.28	.01	.16	.43	-.26
Lo Estimat /Geometry	.35	.20	.14	-.36	-.12	-.36	.14
Lo Estimate	.20	-.10	.19	-.03	.26	-.43	-.09
Flat	.04	-.06	-.01	-.03	.00	.02	.04

Table 7. Statistics for School Profiles for Mathematics Subscores

Profile Description	Percent of Schools	Total Score	Avg # Students per School	Standard Deviation of Scores	Fit
Hi Geometry	15	208	61	.42	.66
Lo Geometry	13	221	62	.43	.68
Lo Estimate	5	214	61	.33	.56
Lo Patterns Algebra	6	207	58	.30	.53
Lo Data / Geometry	6	216	64	.34	.56
Lo Patterns /Geometry	6	216	60	.41	.68
Lo Data / Patterns	6	210	61	.27	.47
Lo Number /Measure	5	212	69	.31	.56
Lo Number / Estimate	6	217	66	.31	.57
Lo Number /Estim/Prob	5	209	49	.36	.64
Lo Data	5	207	51	.34	.61
Lo Data / Number	6	211	56	.33	.56
Lo Estimat /Geometry	5	217	52	.36	.59
Lo Estimate	3	217	56	.31	.52
Flat	7	214	73	.17	.41

Table 8. Correlations* between School Measures and Profile Fit
(Reading above diagonal, Mathematics below diagonal)

	Reading (N=895)			
Mathematics	Number of Students	Standard Deviation	Fit Index	Total Score
(N=877) Number of Students	-	-.36	-.29	.10
Standard Deviation	-.38	-	.76	-.11
Fit Index	-.41	.92	-	-.13
Total Score	.09	.07	.06	-

* All correlations statistically significant at the .01 level.